Network Simulation For AI Applications

Gleb Mishchenko

@Large Research

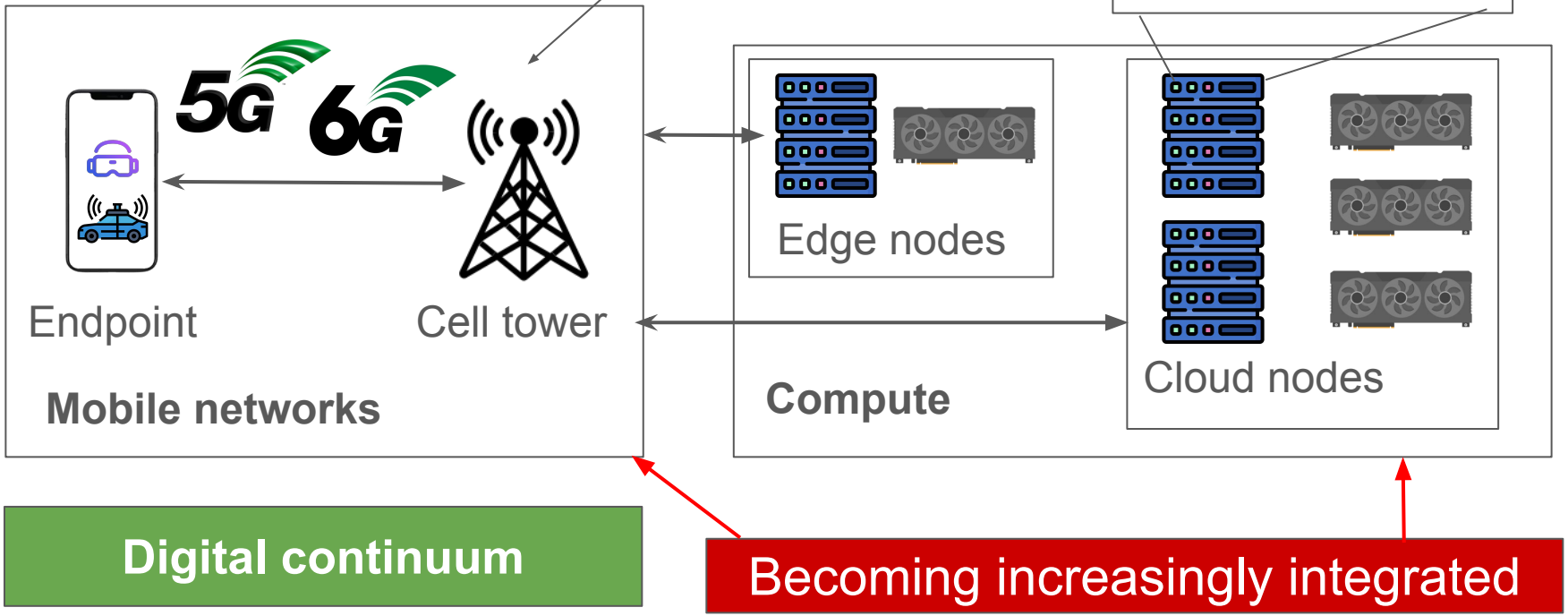
@VU Amsterdam

1st supervisor: Jesse Donkervliet

fzovpec

Context

6G: 100 Gbps bandwidth;
< 0.1 ms latency [1]

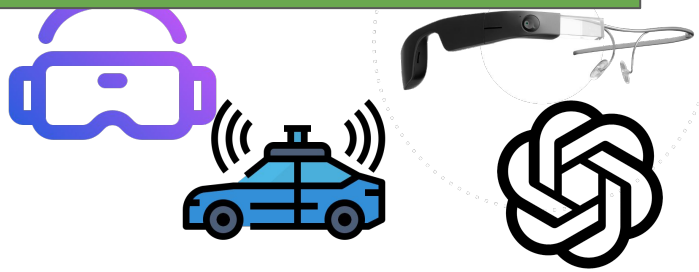


[1] Banafaa, M., Shayea, I., Din, J., Azmi, M. H., Alashbi, A., Daradkeh, Y. I., & Alhammadi, A. (2023). 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. Alexandria Engineering Journal, 64, 245-274.

Problem statement

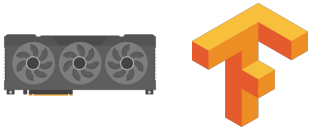
No instrument to analyze trade-offs and make informed decisions

Digital continuum



Are those **feasible**?

Compute latency

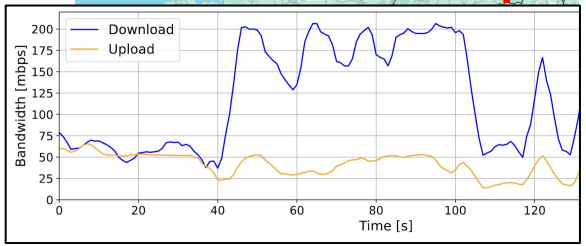


Network latency:

1. technology

2. offloading location

3. jitter

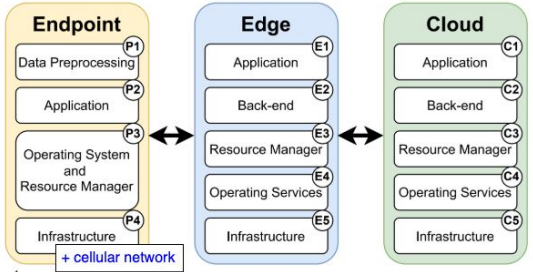


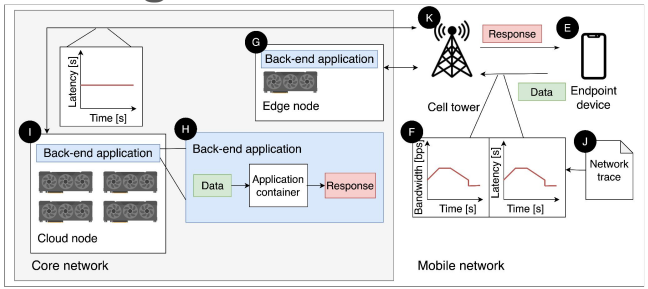
6G TESTBED

Contributions

C1

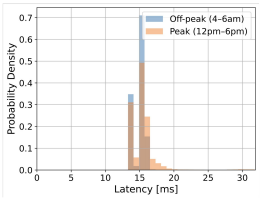
Design and **implementation** of **generalizable** 6G testbed

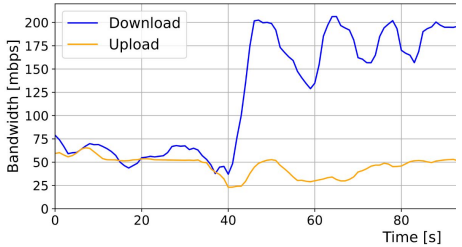




C2

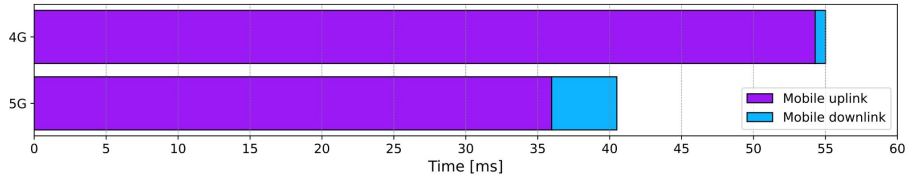
Methodology, **tracing** and **analysis** of state-of-practice Dutch networking



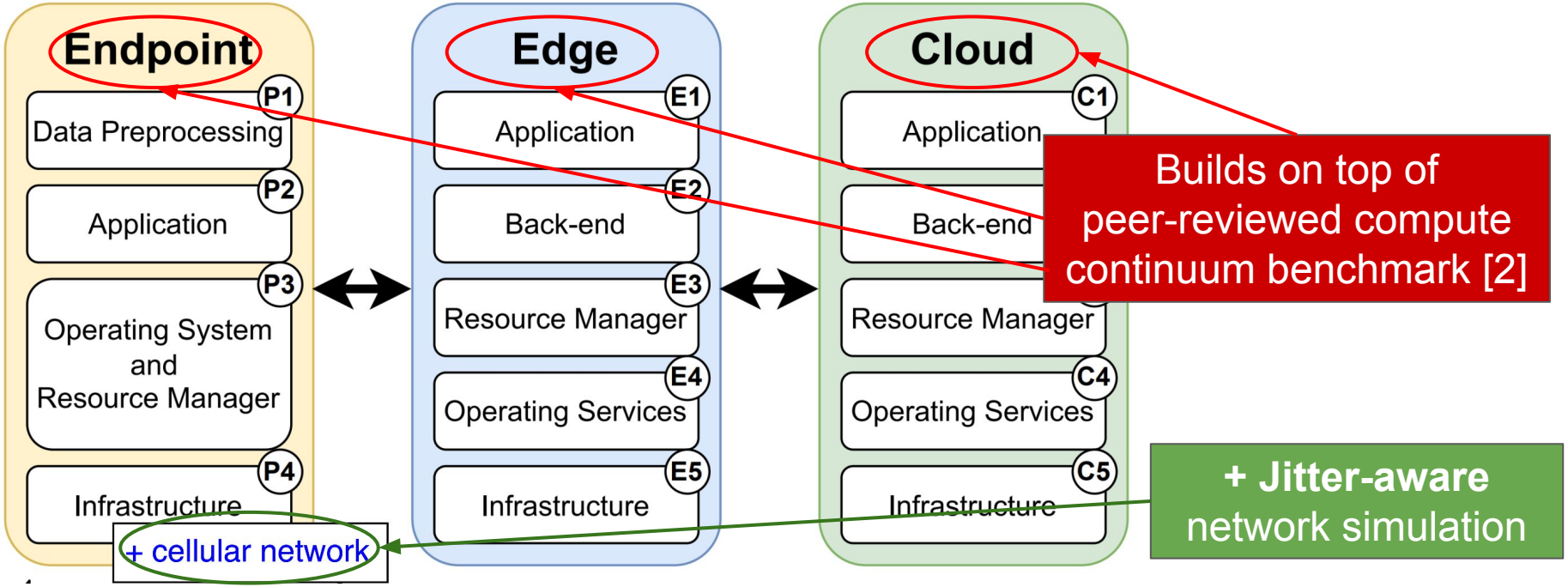


C3

Evaluation of state-of-practice Dutch infrastructure **6G** use-case: **AI-workloads**

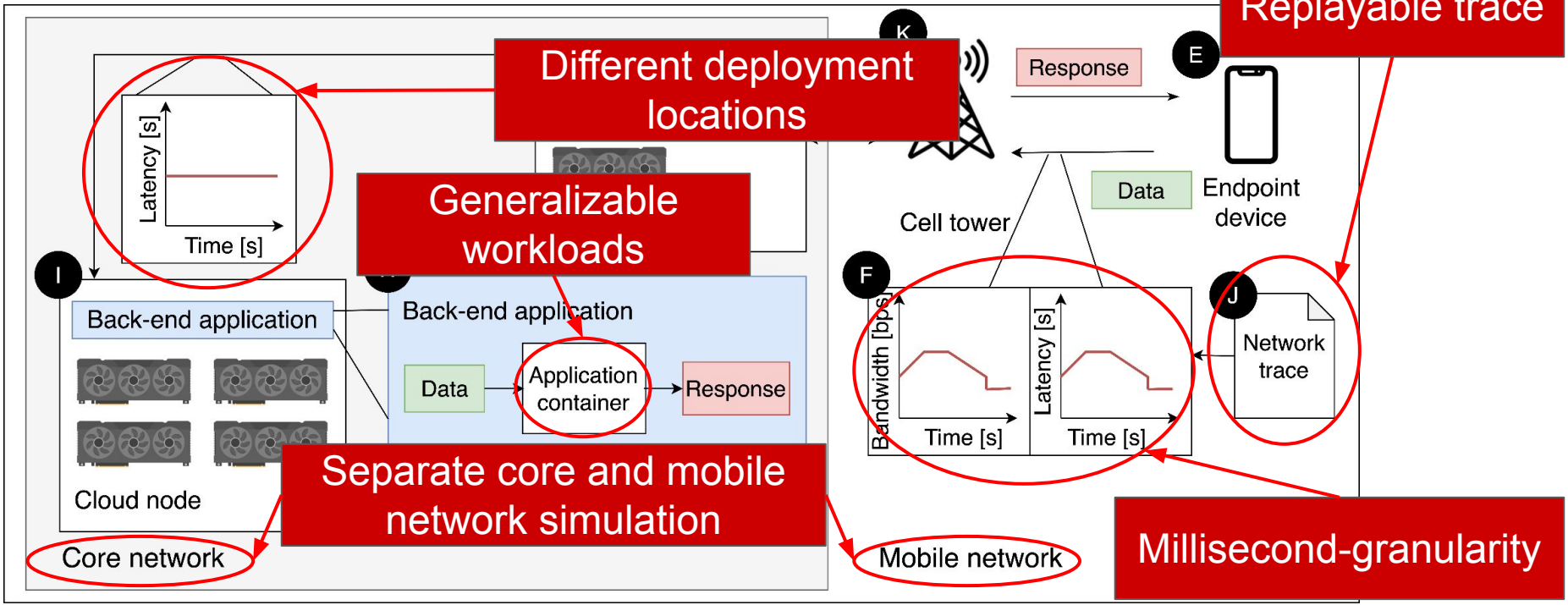


[C1] 6G testbed design and implementation



[2] Jansen, Matthijs, et al. "Continuum: Automate infrastructure deployment and benchmarking in the compute continuum." Companion of the 2023 ACM/SPEC International Conference on Performance Engineering. 2023.

[C1] 6G testbed design and implementation



[C2] Digital continuum performance characterization

~ Cellular network generations

~ Different conditions (obstacles)

Standard	Condition
4G	Outside
5G obst.	Inside*
5G	Outside

* window acts an obstacle between endpoint and the cell tower

Mobile network **traces**

~ Edge/cloud deployments

Recorded over a course of multiple days

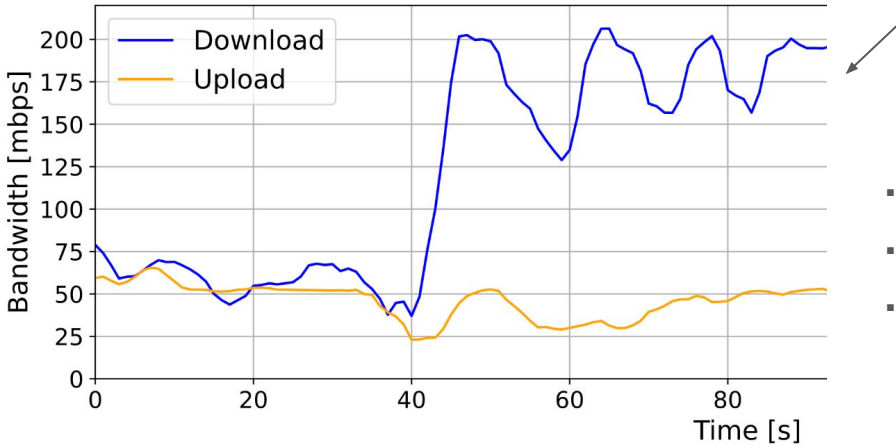
Condition	RTT range [3]
Edge	< 1ms
Local cloud	< 10 ms
Regional cloud	< 20 ms
Transcontinental cloud	< 100 ms

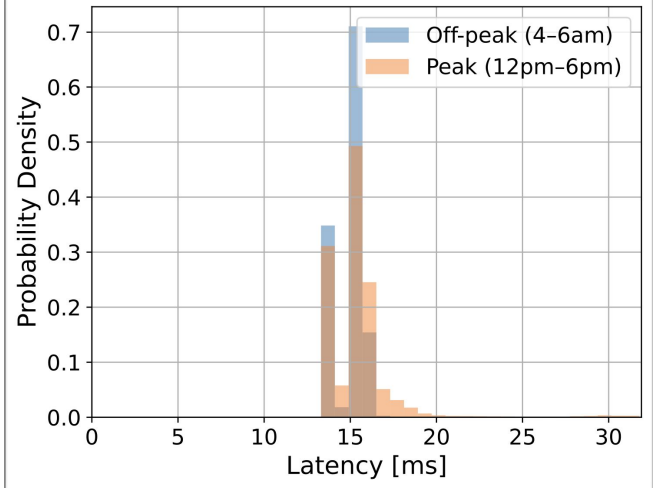
Core network **traces**

[3] Ali-Eldin, A., Wang, B., & Shenoy, P. (2021, November). The hidden cost of the edge: a performance comparison of edge and cloud latencies. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 1-12).

[C2] Digital continuum performance

Cellular networks are highly limited in **upload** capacity





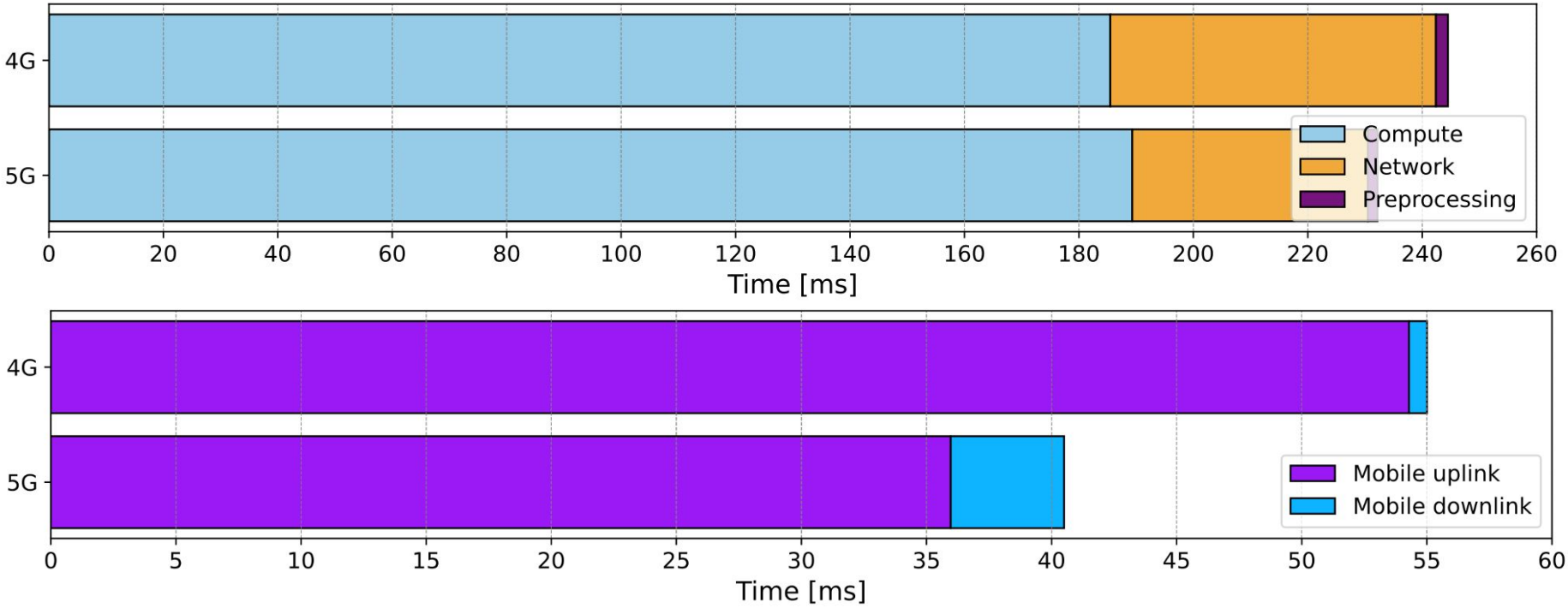
KPN 5G. Amsterdam, The Netherlands

Mobile network performance

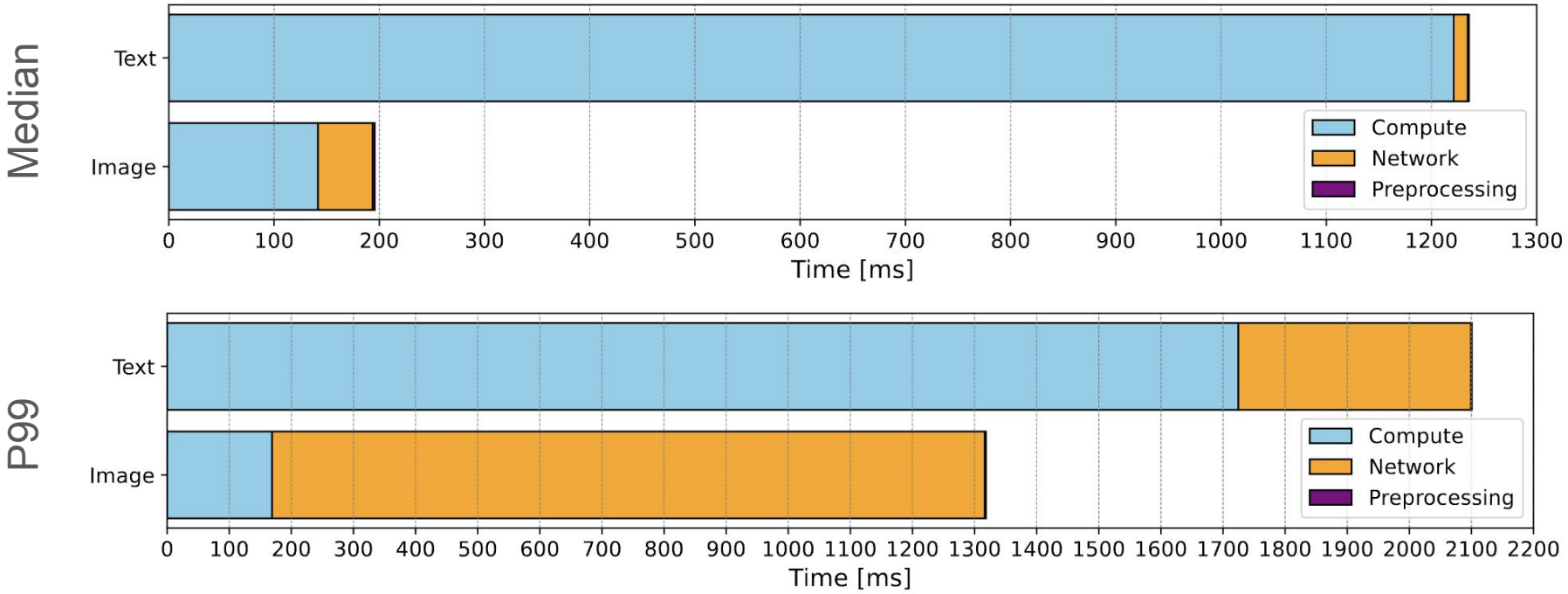
AWS EU-West-1. Paris, France

Core network performance

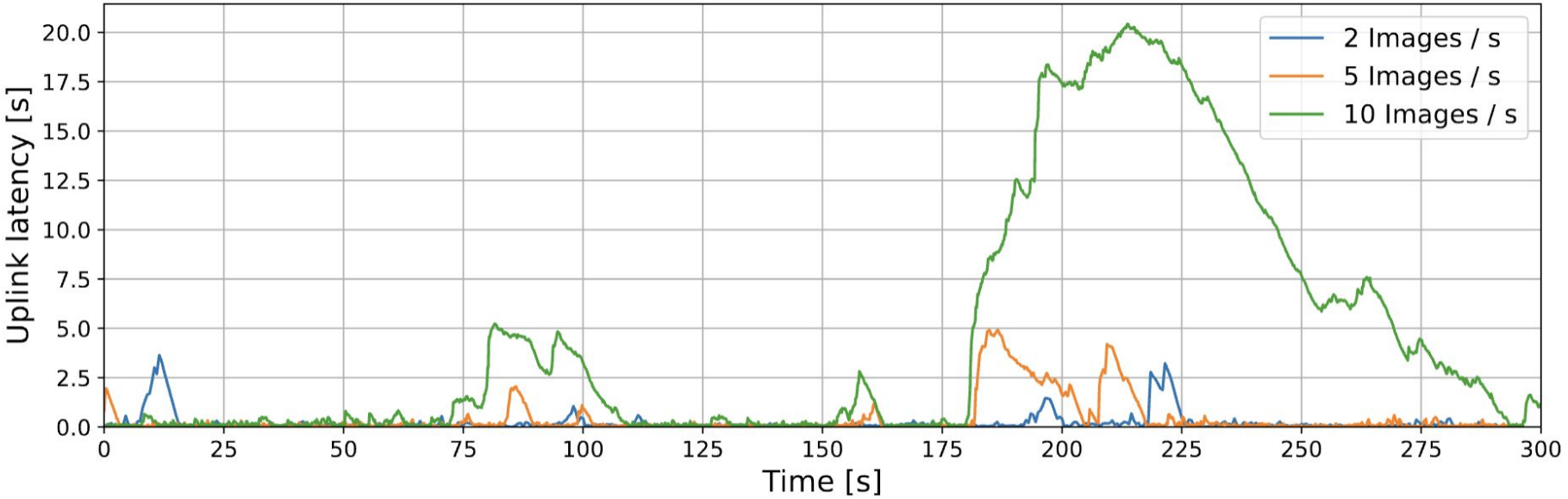
[C3] Impact of cellular network generation



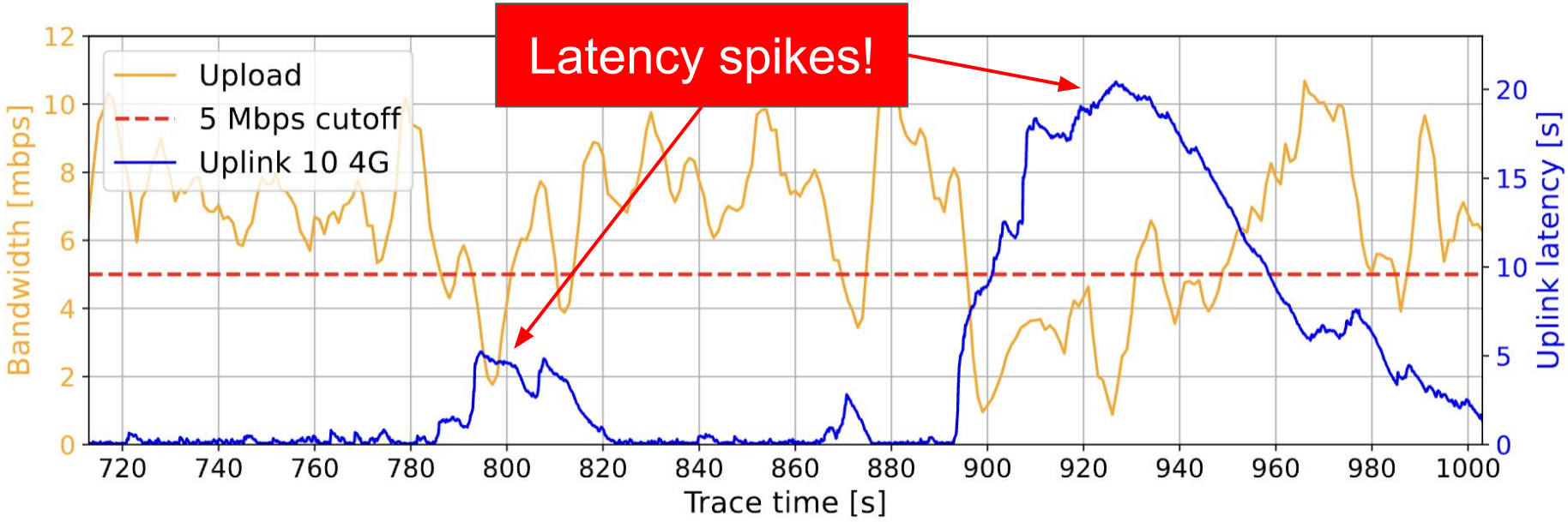
[C3] Impact of workload



[RQ4 PT4] How well does application scale under load?



[C3] Performance under load

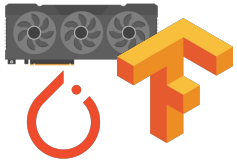


10 requests / second

Conclusion

Design and **implement**
6G testbed

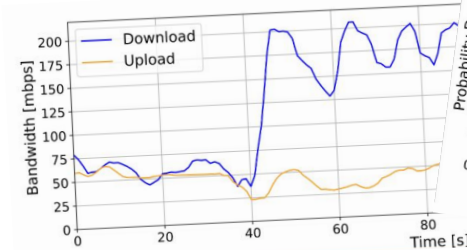
Compute

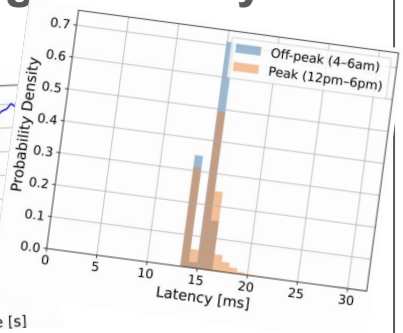


Network

Exploration of
infrastructure trade-offs!

Methodology, **tracing** and **analysis**
network performance





Evaluate state-of-practice networking